DOCUMENT RESUME

ED 157 902 TM 007 098

AUTHOR Harms, Robert A.
TITLE The Development, Validation and Application of an External Criterion Measure of Achievement Test Item Bias.
PUB DATE 28 Mar 78
NOTE 30p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Toronto, Ontario, Canada, March 1978); Bibliography marginally legible

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Achievement Tests; *Content Analysis; *Culture Free Tests; *Evaluation Criteria; *Item Analysis; Measurement Goals; Reading Achievement; Senior High Schools; *Sex Fairness; Stereotypes; *Test Bias; Test Items

ABSTRACT

Based on John Rawls' theory of justice as fairness, a nine-item rating scale was developed to serve as a criterion in studies of test item bias. Two principles underlie the scale: (1) Within a defined usage, test items should not affect students so that they are unable to do as well as their abilities would indicate; and (2) within the domain of a defined usage for any given test item, all students should have equality of opportunity to respond. A group of 316 twelfth-grade students rated a subset of 38 items of a reading test plus five contrived (biased) items. From this group, 247 students identified four or more of the five contrived items as biased. For the reading test, 8 of the 38 items were rated as biased by 30 to 50 percent of the 247 students. Explanations are given for the ratings of these eight items. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED157902

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Development, Validation and Application of an External Criterion Measure of Achievement Test Item Bias

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert A. Harms

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Robert A. Harms

Rt. 2 Box 575

Thonotosassa, Florida, 33592

IM007 098

Running head: Item Bias Criterion Measure

Paper prepared for presentation at the Annual Meeting of the National Council on Measurement in Education Toronto, Canada, March 28, 1978

ABSTRACT

This paper outlines the need for, basis, development, validation and application of an external criterion measure of test item bias.

The criterion measure developed is based on the sole and generally accepted presupposition that fair test items are test items that are just. From this initial presupposition, a criterion measure of item bias based on an operationalization of Rawls' theory of justice, justice as fairness is developed, validated and applied by student raters to items selected from the 1972 TASK II A. Finally, the by-item results and implications of this application are included and discussed.

INTRODUCTORY NOTES

This paper represents an outline of a yet uncompleted research project conducted over the past one and one half years. Included in this overview are components developed on the basis of several disciplines including curriculum, educational measurement and Philosophy. Given the constraints in summarizing and presenting such a paper, much of the detail and/or development rigor is regrettably lost. The author, in recognizing this, welcomes dialogue or requests for clarification on the relevant issues and will provide copies of component parts of the entire document on request. However, the working bibliography is reproduced in full.

The author is and remains indebted to Professor Richard M. Jaeger of the University of North Carolina at Greensboro for his continuing guidance and expertise in the pursuit of this research.

The Development, Validation and Application of an External Criterion Measure of Achievement Test Item Bias

Background and Need for a Criterion Measure

Test item bias, as a deterent to the valid measurement of achievement is well known and documented (Eels et.al., 1952; Merz, 1976). However, despite apparent consensus on its existence and import, little agreement may be seen on the methodology whereby such bias may be detected. Currently, the research literature discloses numerous analytical item bias estimation procedures (well catalogued and outlined by Merz, 1976 and Rudner, 1976), but it is clear that these procedures lack a common conceptual basis. Moreover, in application, these procedures posit desperate results when applied to a common set of items and responses (e.g. Rudner, 1976). Additionally, as each item bias estimation procedure is based on technique-specific assumptions and non-shared methodologies, direct comparisons across procedures remain limited.

The foregoing considerations substantiate the need for the development of an external criterion measure of test item bias. The need for and desirability of such a criterion measure is well documented (e.g. Potthoff, 1966, P. 83; Echternacht, 1974, P. 271;

Green, 1976, P. 3), yet its development has been limited due to the difficulty in obtaining and/or developing and validating such a criterion (Merz, 1976). Although notable methodology studies have advanced criterion measures based on equated groups or stipulated definitions of bias, their results must be considered inconclusive because of the need to accept the assumptions of any such measure.

Background and Basis of the Criterion Measure

The criterion measure that has been developed and applied is based on the sole assumption that if a test item is fair it is just. Therefore, test items that are fair must be considered manifestations and consistent operationalizations of a defensible theory of justice. A comprehensive review of the literature on justice disclosed three dominant themes in oxcidental thought: Plato and Platonic conceptions of justice as outlined in The Dialogues, Utilitarian concepts of justice as posited by Bentham in the latter 18th century, and a contemporary theory, Justice as Fairness developed by John Rawls between 1958-1971. Through an analysis too lengthy for this vehicle but available from the author, it was evident the Rawls' theory alone had the combination of specificity lacking in Plato and avoided the conceptual flaw of the Principle of Utility, the failing of Utilitarianism. The theory proposed by Rawls is a social contract

theory well accepted by the philosophical community as well as sociologists (e.g. Barry, 1967), economists (e.g. Gordon, 1973), legal theorists (e.g. Hart, 1973, Scanlon, 1973) and political scientist (e.g. Barker, 1975). Numerous critical analyses including those of Fried (1963), Runciman (1966), Wolff (1966), Emmett (1969) and Chapman (1973) show Rawls' theory to have been subject to the type of critical analysis (e.g. validation) required from both inside and outside the field of philosophy. Conceptually, justice as fairness is heirarchical, lexiographical and based on two principles:

I. Each person is to have an equal right to the most extensive liberty compatible with a similar liberty for others.

II. Economic and social inequalities are to be arranged so that they are:

   a. reasonably expected to be to everyone's advantage

   b. attached to positions and offices open to all

Principle I states that within any social contract, all contractants are to have the maximum freedom possible, compatible with a similar amount for others. Principle II states that the primary social goods of the society should be open to all with those in greatest need receiving greatest benefit, subject to the constraint that the position of no person/group is denegrated. Operationally stated, Principle I expresses the need for freedom

to be maximized and II, that equality of opportunity be assured.

The Criterion Measure

The criterion measure employed in this research was developed on the basis of these two principles. The measure was designed to assess achievement test bias and was developed to be directly applied to such items by students similar to those who might actually encounter the items. Consistent with Rawls' principles and the discussion in the previous section, Principle I states that an examinee should be free to answer any test item and Principle II requires that all examinees have an equal opportunity to answer. It should be stressed that"answering" is to be considered synonymous with "responding" and that ability to correctly answer the question is no of issue--only that the examinee be free to answer and have an equal opportunity to do so. Therefore, for an item to be fair (non-biased) for any well-defined application, all students or groups of students must be "free" psychologically and emotionally to respond and have an equal opportunity to do so. For application by twelfth grade students, these two principles were rephrased:

Principle I

> Within a defined usage (contract) test items should not "turn-off" students so that they are unable to do as well as their abilities would indicate.

Principle II

Within the domain of a defined usage (contract) for any given test item, all students should have equality of opportunity to respond.

A review of literature on item bias, test item properties, test taking set, minority group relations and theoretical conceptions of bias posited the following nine criteria. Criteria A-E are germane to Principle I whereas Criteria F-I are related to Principle II. These principles and criteria are:

Principle I

Criteria

A An item must not contain any information that could be offensive to a student's religion or culture.

B An item must be non-sexual in the sense that it must not be designed to offend either sexual group.

C An item must not include depictions of any group that are degrading or stereotypic.

D An item must not portray groups as unequal in ability.

E An item must not, by content or form, cause students to be "turned-off" so that they may not do as well as they are able.

Principle II

Criteria

F The content of the test item must not reflect information and/or skills that may not be expected to be within the educational background of all students or groups that take the item.

G The content of the item must not contain any clues or information that could be seen to work to the benefit of any group or either sex.

H The content of the item must be shared to all and not specific to any one group or sex.

I Group specific language, vocabulary or reference pronouns (he, she, etc.) must not be included in items.

The need and background for each of the foregoing criteria are summarized as follows:

Criterion A

The basis of the content validation of this criterion is through an appeal to logic. It appears plausible to conclude that a test item embodying any connotations likely to debase individuals or groups could negatively effect their potential of satisfactory performance on an item. The relationship of this criterion to Rawls' Principle I is similarly evident: such affronts are clearly attempts to detract from an individual's liberty to express themself.

Criterion B

Similar to the preceding, the most cogent appeal is to logic. Materials that are offensive to one group will clearly have a negative effect on cognitive functioning. Test items that favor one sex do so at the expense of the other. Inclusion of this criterion is supported by governmental edicts requiring equality of opportunity and the Association for Measurement and Evaluation in Guidance (AMEG) report on sex bias in measurement (1973).

Criterion C

Inclusion of a criterion related to stereotyping of groups is important to the overall description of test item bias. Stereotypic depictions of groups such as women being less capable than men (Faggen-Steckler, McCarthey and Tittle, 1974) is an example of a basic injustice long seen as erroneous. Although the possible long term effects of such depictions are now coming under investigation (e.g. Anastasi, 1973) the possible suppressive effect on test item performance is patent. Clearly, such depictions which serve only to incorrectly and pejoratively portray individuals or groups are antithetical to Principle I.

Criterion D

The logic requiring that this criterion be included is similar to criterion C. Although the data on basic intellectual differences in minority groups is voluminous (e.g., Shuey, 1958), controversial (Jensen, 1969, 1974;

Cronbach, 1975) and unresolved, the injustices of depicting groups as unequal in ability or endowments is apparent. Such depictions may only serve to preclude the maximization of liberty required for satisfaction of Principle I.

Criterion E

The logic for inclusion of this criterion stems from the development of the theory of justice, the criterion measure and the psychometric properties of achievement test items. If a test item is presented in a singular manner or requires response skills not shared across contracting groups, the item works to the detriment of those groups not privy to such skills. Such items, therefore, denigrate the liberty of such groups, which is the overriding consideration of Principle I.

Criterion F

Test items that require that skills are nonshared or group specific, or require information not likely to be within the educational background of all individuals are invalid (Groulund, 1971) and unjust (biased) to groups that do not have benefit of such exposure. The relationship of such nonshared materials to item bias is nearly isomorphic. Additionally, nonshared content effectively precludes equality of opportunity to answer the item (Principle II).

Criterion G

The inclusion of textual cues that may work to the advantage of a group(s) but not others is clearly and logically unjust.

Its biasing effect on item performance is similarly logical and evident.

Criterion H

The logic for inclusion of this criterion is obvious; group-specific content clearly precludes equality of opportunity to answer the item, which is the major tenet of Principle II.

Criterion I

The possible effects on examinee test item performance by the language and vocabulary used in the item is well noted (Faggen-Steckler, McCarthey and Tittle, 1974). The use of such language unfairly disadvantages groups not facile with such language/vocabulary and diminishes their opportunity to answer (Principle II).

What is proposed is that Principles I and II, wholistically, represent a comprehensive and exhaustive theory (criterion) of test item bias. However, as an applied criterion measure, the applicational parameters must be defined prior to implementation. Such parameters include the type of testing, the examinees, the expressed need for the testing and the relevant variables (such as assessment of basic skills; relevancy and/or recency of prior training) within the testing situation.

Validation of the Criterion Measure

The criterion measure was initially assessed on the basis of possible deficiency, contamination, or distortion (Brodgen and

Taylor, 1950) and for its utility within parameters (Astin, 1964). The first three considerations were assessed to be met by a panel of three expert raters facile in the issue of bias. The "applied parameters" were subject to validation both prior and subsequent to the measures' operationalization for use by student raters. The content validity of the measure was initially investigated through an expansion of the foregoing analysis of each criterion. The validity of the two Principles on which Criteria A-I are based was assessed by polling expert (Ph.D. in educational measurement and/or research) raters to determine the correspondence of their definition of test item bias to that implicit in Principles I and II. This analysis showed that 85% of those polled proffered a definition including both I and II and none posited a definition offering either plausable counter-hypotheses to Principles I and II or a definition which did not include either of them.

The validity of the measure as employed was also assessed. In this analysis, a sample of trained student raters applied criteria A-I to a set of deliberately biased items and TASK items a previous application of the measure showed to be unbiased. The results of this application were then compared to those collected from two samples of experts who rated items with and without benefit of the criterion measure. In each comparison, the ratings were substantially in accord. Moreover, measurement experts without benefit of the criterion measure agreed 75% of the time with the reasons why (e.g. applicable criterion or criteria)

items were biased and 95% of the time that the criterion-identified unbiased items were unbiased.

Finally, the instructional module by which twelfth grade students were instructed in the use of the criterion measure was validated. This instructional component was developed on the basis of the Baker and Schutz instructional product development method (Baker and Schutz, 1975) and in conjunction with subject area instructional specialists. Further, all initial assumptions made on the types of students likely to be encountered were validated by a separate data collection and the instruction was field-tested through three iterations including a participant-observer evaluation.

Preliminary Application

The preliminary application of the criterion measure was completed in the winter of 1977-78. In this application, 316 Florida twelfth grade students were instructed in the use of the criterion measure. A subset of 38 items was selected from the Reading section of the 1972 TASK level II A. One of each of three item types, reading, comprehension, vocabulary and sentence completion was selected and given to each student rater. Additionally, raters were also given the five "contrived" or deliberately biased items as a method of assessing the validity of their ratings. Ratings of students were treated as nominal-dichotomous (biased or non-biased) data. Any students' item ratings were

only used if that student "passed" the cut-off of identifying four of the five contrived items as biased. From this check, 247 student ratings were included in the data analysis. The results of this application are included in Figure 1.

Figure 1

Discussion

Although no definitive guidelines have been established at this time, for purposes of discussion, "high" bias ratings will be defined as those in which greater than 30% of the total respondents identified an item as biased. Therefore, such items were 6, 43, 44, 45, 46, 47, 48 and 66. Their apparent basis for inclusion may be detailed as follows.

In item 6, post-rating interview of students disclosed that many felt the rich/poor dichotomy presented in this reading comprehension question and manifest in item 6 represented content that could be offensive to American Indian respondents. Items 43-45 were part of a single sentence completion question. In this three item group 37%-45% of student raters detected potential bias, noting that the form of the item (phraseology) which included the terms "Everyone knows. . ." and "What many _____ aren't aware of. . ." represented a "put down" of those that didn't know. Such a put-down was interpreted as a turn-off due to content or form, a violation of Criterion E.

Items 46-48 which were a similar three item group were photography-related. These items made student raters feel that those facile in photography would be unfairly advantaged in responding to the question. Such advantagement was interpreted as a violation of Criterion F which holds that the content or skill required to answer should be shared by all respondents. Raters also felt that this item could possibly contain clues that could work to the advantage of photography hobbyists. Such potential clues are a violation of the anti-shared clue provision of Criterion G. The final high bias identification item, 66, was a part of a vocabulary matching item. Post-rating interviews of students made it apparent that the word, which has strong drug-related connotations was seen as a "turn-off" by a high percentage of students.

Implications

Analyzing test items for bias is most useful in the test construction phase when the test is easily amenable to change. The included criterion measure has been systematically applied in several such applications with encouraging results. It appears to this author, however, that the measure is very possibly conservative and its application will lead to the selection of a larger number of items as biased than post-hoc statistical procedures. However, as it is easier to rewrite abberant items

than to speculate on their effect on scores, such conservatism is likely to be both long-term efficient and beneficial.

The second phase of this research will be to employ five of the published analytical bias estimation procedures to (point biserial, factor analysis, Chi-square, Rasch model and delta value) to the TASK items previously examined by the criterion measure. The results of this comparision will be used to provide a comparative concurrent validation of these analytical estimators. What is advocated herein is not a replacement or even an additional item bias estimation procedure, but rather an ongoing test development and item writing tool that may function as a benchmark measure of test item bias.

REFERENCES

Altham, J. Rawls' difference principle Philosophy 1973, 48, 1, 75-78

AMEG commission report on sex bias in interest measurement Measurement and Evaluation in Guidance 1973, 6, 3, 171-177

Anastasi, A. Common falacies about heredity environment and human behavior ACT Research Bulletin No. 58 Iowa City, Iowa: American College Testing Program, 1973

Angoff, W. A technique for the investigation of differences Paper presented at the annual meeting of the American Psychological Association 1972

_____ and Sharon, A. Patterns of test and item difficulty on a test of English as a foreign language Princeton, New Jersey: Educational Testing Service, 1972

_____ and Ford, S. Item-race interaction on a test of scholastic aptitude Journal of Educational Measurement, 1973, 10, 2, 95-106

Arrow, K. Some ordinalist-utilitarian notes on Rawls' theory of justice Journal of Philosophy, 1973, 70, 9, 245-263

Astin, A. Criterion centered research Educational and Psychological Measurement, 1964, 24, 4, 245-263

Bailar, N. Dictionary of sampling terms Mimeo: American Educational Research Association Sampling Institute, Clearwater, Florida, 1975

Barker, B. Justifying justice: problems of psychology politics and measurement American Political Science Review, 1975, 52, 331-386

Barker, E. Social contract London: Oxford University Press, 1966

Barry, B. John Rawls and the priority of liberty Philosophy and Public Affairs, 1973, 2, 3, 274-290

Binkley, J. Contemporary ethical theories New York: Citadel Press, 1961

Birnbaum, A. Some latent trait models and their use in inferring an examinees ability In F. Lord and M. Novick Statistical theories of mental test scores Reading, Ma.: Addison-Wesley, 1968.

Blumrosen, A. Strangers in paradise: Griggs vs. Duke Power Company and the concept of employment discrimination Michigan Law Review, 1972, 71, 59-110

Breland, H. and Ironson, H. De Funis reconsidered: A comprehensive analysis of alternative admissions Journal of Educational Measurement, 1976, 13, 1, 89-99

Brock, D. Recent work in Utilitarianism American Philosophical Quarterly, 1973, 10, 4, 241-269

Brogden, H. and Taylor, E. The theory and classification of criterion bias Educational and Psychological Measurement, 1950, 10, 159-186

Brown, D. John Rawls: John Mill Dialogue, 1973, 12, Summer, 477-479

Campbell, D. and Fisk, D. Convergent and discriminent validation by the multi-trait multi-method matrix Psychological Bulletin, 1959 56, 81-105

Cardell, C. and Coffman, W. A method of comparing the performance of different groups on the items in a test (ETS RB 64-61) Princeton, New Jersey: Educational Testing Service, 1964

Canady, H. G. The psychology of the negro In P.L. Harriman (ed.) The encyclopedia of psychology New York: Philosophical Library, 1946

Caponigni, A. A history of western philosophy Notre Dame: Notre Dame Press, 1971

Chapman, J. Justice and fairness Nomos VI, 1963

Choptainy, L. A critique of John Rawls' principles of justice Ethics, 1973, 83, 1, 146-150

Cleary, T. Test bias: the prediction of grades of negro and white students at integrated colleges Journal of Educational Measurement, 1968, 5, 115-124

_____ and Hilton, T. An investigation of item bias Educational and Psychological Measurement, 1968, 28, 61-75

_____ ;Humphries, L.; Kendrick, S.; Wesman, A. Educational uses of tests with disadvantaged students American Psychologist, 1975, 30, 15-40

Cohen, M. A theory of justice New York Times Book Review Section 7/16/72

Cole, N. Bias in selection Journal of Educational Measurement, 1973, 10, 237-255

Cornford, F. The republic of Plato New York: Oxford University Press, 1966

Cronbach, L. Essentials of psychological testing New York: Harper and Row, 1959

_____ Test validation In R.L. Thorndike (ed.) Educational measurement Washington:American Council on Education, 1971

_____ Five decades of controversy over mental testing American Psychologist 1975,30, 1-14

_____ Equality in selection, where psychometrics and political philosophy meet Journal of Educational Measurement, 1976, 13, 1, 31-42

_____; Nanda, H. and Rajaratnam, N. The dependability of behavioral measurements: Multifacet studies of generalizability New York: Wiley, 1972

Diamond, E. Minimizing sex bias in testing Measurement and Evaluation in Guidance, 1976, 9, 1, 28-33

Daniels, N. Reading Rawls New York: Basic Books, 1973

Darlington, R. Another look at culture fairness Journal of Educational Measurement, 1971, 8, 71-82

_____ A defense of "rational" personnel selection and two methods Journal of Educational Measurement,1976,13,1

Durant, W. The story of philosophy New York: Harcourt , 1963

Dworkin,R. The original position University of Chicago Law Review, 1973, 40, 3, 500-533

Eagle, N. and Harris, A. Interaction of race and test on reading performance scores Journal of Educational Measurement, 1969, 6, 3, 131-135

Echternacht, G. A quick method of determining test bias Educational and Psychological Measurement 1974, 34, 271-280

Eels, K.;Davis, A.; Havighurst, R.; Herrick, V.; Tyler, R. Intelligence and cultural differences Chicago: University of Chicago Press, 1952

Einhorn, H. and Bass, A. Methodological considerations relevant to discrimination in employment testing Psychological Bulletin, 1971, 75, 261-269

Emmett, D. Justice Proceedings of the Aristotelian Society 1969, 72, 1, 238-246

English, H. and English, A. A comprehensive dictionary of psychological terms New York: Longmans, Green, 1958

Faggen-Steckler, J.; Mc Carthey, K. and Tittle, C. A quantative method of determining sex bias in standardized tests Journal of Educational Measurement, 1974, 11, 3, 151-163

Field, W. Plato Oxford: Oxford University Press, 1949

Feinberg, J. Duty and obligation in the non-ideal world Journal of Philosophy, 1973, 70, 9, 263-275

Fisk, M. History and reason in Rawls'theory In N. Daniels (ed.) Reading Rawls New York: Basic Books, 1973

Fleming, M. The anchor test study: The long and the short of it Journal of Educational Measurement 1975, 12, 3, 201-213

Glass, G. and Stanley, J. Statistical methods in education and psychology Englewood Cliffs: Prentice Hall, 1970

Gough, J. The social contract Oxford: Clarendon Press, 1963

Gorden, S. John Rawls's difference principle, utilitarianism and the optimum degree of inequality Journal of Philosophy , 1973, 70, 9, 275-280

Goslin, D. Standardized ability tests and testing Science, 1968, 159, 851-855

Grant, D. and Bray, D. Validation of employment tests for telephone company installation and repair occupations Journal of Applied Psychology, 1970 54, 7, 7-14

Green, D. Procedures for assessing bias in achievement tests Paper presented at the NIE conference on test bias, 1975

_____ Racial and ethnic bias in achievement test construction Monterey, Ca. CTB/ Mc Graw Hill, 1971

_____ Reducing bias in achievement tests Paper presented at the annual meeting of the American Educational Research Association, 1976

_____ and Draper, J. Exploratory studies of bias in achievement tests Paper presented at the annual meeting of the American Educational Research Association, 1972

Grounland, N. Measurement and evaluation in teaching New York: Mc Millan, 1971

Hart, H. Rawls on liberty and its priority University of Chicago Law Review, 1973, 40, 534-555

Haskar, V. Rawls' theory of justice Analysis , 1972, 32, 5, 149-152

Hicks, M.; Dolon, T.; Wallmark, M. Sex differences in item responses on the graduate record examination Paper presented at the annual meeting of the National Council on Measurement in Education, 1976

Hodgson, D. Consequences of utilitarianism Oxford: Clarendon Press, 1967

Hoefner, R. and Strickland, G. Investigating test bias Los Angeles: Canter for the Study of Evaluation, UCLA, 1972

Hollman, T. Differential validity: A problem with tests or criteria Paper presented at the Midwest Psychological Association, 1973

Jaeger, R. Matrix sampling Materials developed for American Educational Research Association Sampling Institute Clearwater, Florida, 1975

_____ Personal communication January 10, 1977

Jensen, A. How biased are culture loaded tests ? Genetic Psychology Monographs, 1974, 90, 185- 244

_____ How much can we boost I.Q. and scholastic achievement ? Harvard Educational Review, 1969, 39, 1, 1- 123

_____ Test bias and construct validity Phi Delta Kappan, December, 1976

Kallingal, A. The prediction of grades for black and white students at Michigan State University Journal of Educational Measurement, 1971, 8, 263-265

Klineberg, O. Race and selective migration New York: Columbia University Press, 1936

Lesnoff, M. John Rawls' theory of justice Political Studies, 1971, 19, 1, 63-80

Linn, R. Fair test use in selection Review of Educational Research, 1973, 43, 139-161

_____ In search of fair selection procedures Journal of Educational Measurement, 1976, 13, 1, 53-58

_____ and Werts, C. Considerations for studies of test bias Journal of Educational Measurement, 1971, 8, 1, 1-4

Lord. F. A theory of test scores Psychometric Monograph No. 7 Princeton, New Jersey: Educational Testing Service, 1952

Lyons, D. Rawls vs. utilitarianism The Journal of Philosophy, 1972, 79, 18, 535-545

Maccoby, E. Sex differences in intellectual functioning In E. E. Maccoby (ed.) The development of sex differences Stanford: Stanford University Press, 1966

Mandlebaum, M. A theory of justice History and Theory 1973, 12, 2, 240-250

Merz, W. Factor analysis as a technique for analyzing test bias Paper presented at the annual meeting of the California Educational Research Association, 1973

_____ A biased test may be fair, but then what does bias really mean ? Paper presented at the annual meeting of the California Educational Research Association

_____ Estimating bias in test items utelizing principle components analysis and the general linear models solution Paper presented at the annual meeting of the American Educational Research Association, 1976a

_____ Test fairness and test bias: A review of procedures Paper presented at the Office of Education Invitational Conference on Achievement Testing of Disadvantaged and Minority Students for Educational Program Evaluation 1976 b

Moore, G. Ethics London : Oxford University Press, 1903

Nagel, T. Rawls on justice Philosophical Review, 1973, 82, 4, 220-234

Oakland, T. Assessing minority group children: challenges for school psychologists Journal of School Psychology, 1973, 11, 4, 294-302

O'Conner, J. Wolff, Rawls and the principles of justice Philosophical Studies 1973, 19, 6, 93-95

Ozenne, D.; Van Gelder, N. and Cohen, A. Emergency school aid act national evaluation and achievement test restandardization Santa Monica, Ca. :System Development Corperation, 1974

Peterson, N. and Novick, M. An evaluation of some models of culture fair selection Journal of Educational Measurement, 1976, 13, 1, 3-30

Pfeifer, C. and Sedlack, W. The validity of academic predictions for black and white students at predominately white universities Journal of Educational Measurement, 1971, 8, 253-261

Pollack, L. A dilemmafor Rawls ? Philosophical Studies, 1971, 22, 4, 262-265

Potthoff, R. Statistical aspects of the problem of biases in psychological tests Institute of Statistics Mimeo Series No. 479 Chapel Hill: University of North Carolina, Department of Statistics, 1966

Rawls, J. Symposium: justice as fairness Journal of Philosophy, 1957, 52, 22, 653-670

_____ Justice as fairness Philosophical Review 1958, 67, 2, 164-194

_____ A theory of justice Cambridge: Harvard University Press, 1971

Rudner, L. Approaches toward bias item identification Washington: Gallaudet College, 1976

_____ Item and foremat bias and appropriateness Washington: Gallaudet College, 1976 b

_____ Personal communication December 12, 1976 c

_____ Personal communication December 28, 1976 d

Runciman, W. Relative deprivation and social justice Berkeley: University of California Press, 1966

Samuda, R. Psychological testing of American minorities New York: Dodd and Mead, 1975

Scanlon, T. Rawls' theory of justice University of Pennsylvania Law Review, 1973, 121, 5, 1020-1099

_____ Rawls' theory of justice In N. Daniels Reading Rawls New York: Basic Books, 1973 b

Scheffe, H. The analysis of variance New York:Wiley, 1959

Schmidt, F. and Hunter, J. Racial and ethnic bias in psychological tests American Psychologist, 1974, 72, 1-8

Schueneman, J. A procedure for evaluating item bias in the absence of an outside criterion Paper presented at the annual meeting of the American Educational Research Association, 1976

_____ Remarks in symposium 15.12: The assessment of bias in the absence of an outside criterion Annual meeting of the American Educational Research Association, 1975

Scott, G. John Rawls' difference principle, utilitarianism and the optimum degree of inequality Journal of Philosophy, 1973, 70, 4, 263-275

Sen, A. Rawls vs. Bentham: an axiomatic examination of the pure distribution problem Theory and Decision, 1973, 4, 3, 301-310

Shuey, A. The testing of negro intelligence New York: Social Science Press, 1958

Sirotnik, K. Introduction to matrix sampling in W. J. Popham (ed.) Evaluation in education Berkeley, Mc Cutchen Publishing Co., 1974

Strassberg-Rosenberg, B. and Dolon, T. Context influences on sex differences in performance and aptitude tests Paper presented at the annual meeting of the National Council on Measurement in Education, 1975

Thorndike, R. Concepts of culture fairness Journal of Educational Measurement, 1971, 8, 63-70

_____ and Hagan, E. Measurement and evaluation in education and psychology New York: Wiley, 1961

London Times Literary suppliment London, July 6, 1972

Tittle, C. Fairness in educational achievement testing Education and Urban Society, 1975, 92. 86-103

Tocco, T. Personal communication December 31, 1976

Tyler, L. The psychology of human differences New York: Century Appleton Crofts, 1965

Urry, V. A monte carlo investigation of logistic mental test models Unpublished Doctoral Dissertation Purdue University, 1970

Veale, J. Cultural validity of items and tests: a new approach SCORE Technical Report Iowa City: Westinghouse Learning Corperation, 1975

_____ and Foreman, D. Cultural variation in criterion referenced tests Paper presented at the annual meeting of the American Educational Research Association, 1976

Wesman, A. Writing the test item In R. Thorndike (ed.) Educational Measurement Washington: American Council on Education, 1971

Wolff, R. A refutation of Rawls theorem Journal of Philosophy , 1966, 73, 7, 179-190

Wright, B.; Mead, R. and Draba, R. Detecting and correcting test item bias with a logistic response model Research Memorandum No. 22 Statistical Labratory, Department of Education, The University of Chicago, 1976

Figure 1
Plot of Item (As Biased) Responses
247 Raters